

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-331827

(43) 公開日 平成11年(1999)11月30日

(51) Int.Cl.<sup>8</sup>  
H 0 4 N 7/18  
G 0 6 T 1/00  
H 0 4 N 5/225  
5/232

識別記号

F I  
H 0 4 N 7/18  
5/225  
5/232  
5/268  
C  
F  
D  
C

審査請求 未請求 請求項の数 9 O L (全 15 頁) 最終頁に続く

(21) 出願番号 特願平10-128836

(22) 出願日 平成10年(1998) 5月12日

(71) 出願人 000003223

富士通株式会社

神奈川県川崎市中原区上小田中4丁目1番  
1号

(72) 発明者 枝並 隆文

神奈川県川崎市中原区上小田中4丁目1番  
1号 富士通株式会社内

(74) 代理人 弁理士 柏谷 昭司 (外2名)

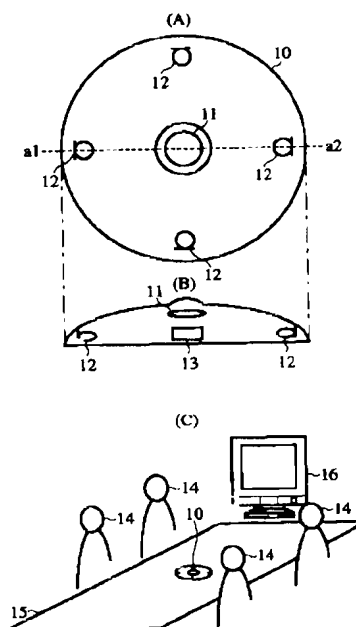
(54) 【発明の名称】 テレビカメラ装置

(57) 【要約】

【課題】 魚眼又は超広角レンズ及び可変指向性マイクロフォンを用いたテレビカメラ装置に関し、テレビ会議中に音声が発生されない場合でも人物の追尾を安定して行い、アクティビティのあるテレビ会議の映像を生成し、雑音やエコーを低減した高品質の臨場感のあるテレビ会議を実現し、小型で軽量、且つ可動部の無いテレビカメラ装置を提供する。

【解決手段】 テレビカメラ装置10は、中央部に魚眼又は超広角レンズ部11とCCD撮像部13を備え、周辺部に複数の無指向性マイクロフォン12が配列され、無指向性マイクロフォンの位相制御により可変指向性マイクロフォンとすると共に、音源位置の方向(話者方向)を判定し、該音源位置方向を追尾し、音源位置方向の画像(話者の人物像)を切り出して映像信号を生成する構成を有する。

本発明の実施の形態の  
テレビカメラ装置の構成と使用状況を示す図



**【特許請求の範囲】**

【請求項1】 テレビ会議又は遠隔監視に使用されるテレビカメラ装置において、

魚眼又は超広角レンズのカメラと、

複数の無指向性マイクロフォンを配列した可変指向性マイクロフォンと、

前記可変指向性マイクロフォンにより、音源位置の方向を判定して、該音源位置方向を追尾する手段と、

前記音源位置方向の画面を切り出して映像信号を生成する手段とを備えたことを特徴とするテレビカメラ装置。

【請求項2】 テレビ会議又は遠隔監視に使用されるテレビカメラ装置において、

魚眼又は超広角レンズのカメラと、

複数の無指向性マイクロフォンを配列した可変指向性マイクロフォンと、

前記可変指向性マイクロフォンにより、音源位置の方向を判定して、該音源位置方向を追尾する手段と、

前記魚眼レンズ又は超広角レンズのカメラで撮像した画像のフレーム間差分により移動像を検出して追尾する手段と、

前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とのいずれか一方を選択して画面を切り出し、映像信号を生成する手段とを備えたことを特徴とするテレビカメラ装置。

【請求項3】 前記音源位置方向を追尾する手段は、有声音を検出する手段を有し、有声音が検出される音源位置方向の画面を切り出し、映像信号を生成する構成を有することを特徴とする請求項1又は2記載のテレビカメラ装置。

【請求項4】 前記音源位置方向を追尾する手段は、有声音が検出されない音源位置方向の利得を低下させるように前記無指向性マイクロフォンの出力音信号を合成する構成を有することを特徴とする請求項1乃至3いずれかに記載のテレビカメラ装置。

【請求項5】 前記音源位置方向を追尾する手段は、モニタテレビ装置のスピーカー音信号を検出する手段を有し、前記無指向性マイクロフォンの出力音信号と該モニタテレビ装置のスピーカー音信号との相関を計算して、スピーカー位置を判定し、前記無指向性マイクロフォンから出力されるスピーカー音を抑圧するように、前記無指向性マイクロフォンの出力音信号を合成する構成を有することを特徴とする請求項1乃至4いずれかに記載のテレビカメラ装置。

【請求項6】 前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とを備えたテレビカメラ装置において、音源位置方向を追尾する手段を優先させて画面を切り出し、話者を強調して表示する映像信号を生成する構成を有することを特徴とする請求項2乃至5いずれかに記載のテレビカメラ装置。

【請求項7】 前記音源位置方向を追尾する手段と移動

像を検出して追尾する手段とを備えたテレビカメラ装置において、移動像を検出して追尾する手段を優先させて画面を切り出し、移動人物を強調して表示する映像信号を生成する構成を有することを特徴とする請求項2乃至5いずれかに記載のテレビカメラ装置。

【請求項8】 前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とを備え、その両者の追尾手段のいずれかをキー入力により選択する構成を有することを特徴とする請求項2乃至5いずれかに記載のテレビカメラ装置。

【請求項9】 魚眼又は超広角レンズのカメラと複数の無指向性マイクロフォンとを一体化し、前記魚眼又は超広角レンズを中央部に配置し、前記複数の無指向性マイクロフォンを周辺部に配列して構成したことを特徴とする請求項1乃至8いずれかに記載のテレビカメラ装置。

**【発明の詳細な説明】****【0001】**

【発明の属する技術分野】本発明は、魚眼又は超広角レンズ及び可変指向性マイクロフォンを用いたテレビカメラ装置に関し、特に、テレビ会議用のカメラ装置又は遠隔監視用のモニタカメラ装置等に使用され、画像信号から人物等の移動像とその方向を識別すると共に、音声信号から発言者の方向を識別し、魚眼又は超広角レンズで撮影した映像の中から、移動人物又は発言者の映像を切り出して映し出すテレビカメラ装置に関する。

**【0002】**

【従来の技術】従来のテレビカメラ装置は、固定カメラ又はリモコン操作式の首振りカメラとマイクロフォンとにより、会議参加者の映像と音声とを取り込み、その映像信号と音声信号とを遠隔の会議場のモニタテレビ装置に回線網を介して送出する。

【0003】前述の固定カメラによる映像は、撮像エリアが一定で動きがなく、そのため映像画面が単調であり、又、会議参加者が撮像エリアから移動すると、モニタテレビ画面から外へはみ出してしまい、固定カメラによるテレビ会議装置は、円滑なテレビ会議には不向きである。一方、リモコン操作式の首振りカメラは、会議参加者が、会話をしながらカメラの首振りを操作しなければならず、その操作が煩わしく、そして首振り動作の反応が遅く、操作性及び使用勝手が悪い。

【0004】そこで、発言を行っている話者を自動的に識別して該話者を映し出すよう、カメラの向きを自動的に旋回させて追尾させたり、話者を示す表示を付加して映し出すテレビカメラシステムが考え出されている（例えば、特開平5-153582号公報、特開平7-15710号公報、特開平6-351015号公報、特開平5-207449号公報等参照）。

【0005】図12は従来のテレビ会議用のカメラ旋回システムを示す図である。図の(A)はカメラ旋回システムの構成、図の(B)はその使用状況を示す。同図に

において、12-1はn個の単一指向性マイクロフォンを放射状に配置した第1のマイクロフォン装置、12-2はm個の高感度単一指向性マイクロフォンを放射状に配置した第2のマイクロフォン装置、12-3及び12-6はそれぞれ第1及び第2のマイクロフォン装置の中で、最も大きい音声が入力されたマイクロフォンを検出する音声認識部、12-4は第2のマイクロフォン装置の旋回を制御するマイク制御部、12-5はカメラの旋回を制御するカメラ制御部、12-7はテレビカメラである。

【0006】第1のマイクロフォン装置12-1のn個の単一指向性マイクロフォンは、各々、会議卓の円周をn等分した各円弧の中心方向にその指向性に向けて固定的に配置されている。今、図の(B)に示すように、会議卓の円周部のn等分した或る1つの円弧12-8の方向に指向性に向けたマイクロフォンからの音声入力、最も大きい音声として第1の音声認識部12-3により認識されると、音声認識部12-3は、該円弧12-8の位置方向の情報を、マイク制御部12-4とカメラ制御部12-5とに送出する。

【0007】マイク制御部12-4は、第2のマイクロフォン装置12-2全体を巡回させ、最も大きい音声、認識された前述の円弧部12-8の中心方向に第2のマイクロフォン装置12-2全体を向ける。第2のマイクロフォン装置12-2のm個の高感度単一指向性マイクロフォンは、各々、該円弧部12-8の円周を更にm等分した方向にその指向性に向けて配置され、第2の音声認識装置12-6は、円弧部12-8をm等分したm個の位置方向の中から、最も大きい音声が入力される位置方向12-9を認識し、その位置方向情報をカメラ制御部12-5に送出する。

【0008】カメラ制御部12-5は、第1及び第2の音声認識部から送出される位置方向情報を基に、最も大きい音声が入力される位置方向12-9が話者の方向と判定し、該位置方向12-9に向けてテレビカメラ12-7を巡回させる。

【0009】図13は従来の話者位置を表示するテレビ会議システムを示す図である。同図において、13-1及び13-2は第1及び第2のマイクロフォン、13-3は音声回路、13-4は位相差検出回路、13-5は演算回路、13-6は位置情報回路、13-7はテレビカメラ、13-8は映像回路、13-9は映像合成回路、13-10は映像符号器、13-11はネットワーク、13-12は受信回路、13-13はモニタテレビ、13-14はスピーカーである。

【0010】第1及び第2のマイクロフォン13-1、13-2に入力される音声の位相差を位相差検出回路13-4により検出し、該位相差から話者の位置を演算し、その位置情報から話者位置を表示する映像信号と、テレビカメラ13-7で撮影した映像(映像回路13-

8の出力映像信号)とを、映像合成回路13-9により合成し、該合成した映像信号を符号化し、ネットワーク13-11を介してモニタテレビ13-13に映し出すものである。

【0011】

【発明が解決しようとする課題】従来のテレビ会議用のシステムは、マイクロフォンとテレビカメラの設置位置とを、あらかじめ会議参加者に向けて設定するというプリセット動作が必要であり、会議中に、マイクロフォンの位置を変更することはできない。

【0012】又、マイクロフォンは、テレビ会議に参加する人数分のマイクを必要とする。又、テレビカメラの巡回時に発生する機械音が会話に支障をきたす等の問題点を含んでいる。又、2個のマイクロフォンから入力される音の位相差により、話者の位置を検出する場合、雑音やモニタテレビ装置のスピーカーから送出される音声に対しても、敏感に反応するため、雑音やスピーカー音の伴う実際のテレビ会議の場面での使用には支障が多い。

【0013】又、音声のみにより、追尾方向を決定するため、対象人物が発声をしないと追尾方向が不安定となり、又、会議場に複数の人物が存在する場合に、複数の人物を映し出したりすることができず、常に一人の会議参加者のみを追尾することしかできない。又、テレビカメラの巡回音が、会議参加者に緊張感や違和感を与えてしまう等の問題を抱えている。

【0014】更に、テレビカメラ装置は、遠隔の相手方の会議参加者を映すモニタテレビ装置とその音声を放音するスピーカーと共に、会議室等に設置され、該スピーカーから放音される音声、テレビカメラ装置のマイクロフォンに回り込んでエコーが発生するため、エコーを消去又は抑制する高価なエコーキャンセラ又はエコーサプレッサを組み込む必要があった。

【0015】本発明は、テレビ会議中に音声が発生されない場合でも人物の追尾を安定して行い、自然でアクティビティのあるテレビ会議の映像を生成することを目的とする。又、話者の位置を特定するのみでなく、マイクロフォンに入力される音源が、雑音、人物の音声又はモニタテレビ装置のスピーカー音であるかに応じて、それぞれの音源に対して異なる指向性を与え、高品質の臨場感のあるテレビ会議を実現することを目的とする。更に、会議場等の音声及び画像をすべて収集し、その中から送出する音声及び画像を抽出する小型で軽量、且つ可動部の無いテレビカメラ装置を提供することを目的とする。

【0016】

【課題を解決するための手段】本発明のテレビカメラ装置は、(1)テレビ会議又は遠隔監視に使用されるテレビカメラ装置であって、魚眼又は超広角レンズのカメラと、複数の無指向性マイクロフォンを配列した可変指向

性マイクロフォンと、前記可変指向性マイクロフォンにより、音源位置の方向を判定して、該音源位置方向を追尾する手段と、前記音源位置方向の画面を切り出して映像信号を生成する手段とを備えたものである。

【0017】又、(2) テレビ会議又は遠隔監視に使用されるテレビカメラ装置であって、魚眼又は超広角レンズのカメラと、複数の無指向性マイクロフォンを配列した可変指向性マイクロフォンと、前記可変指向性マイクロフォンにより、音源位置の方向を判定して、該音源位置方向を追尾する手段と、前記魚眼レンズ又は超広角レンズのカメラで撮像した画像のフレーム間差分により移動像を検出して追尾する手段と、前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とのいずれか一方を選択して画面を切り出し、映像信号を生成する手段とを備えたものである。

【0018】又、(3) 前記音源位置方向を追尾する手段は、有声音を検出する手段を有し、有声音が検出される音源位置方向の画面を切り出し、映像信号を生成する構成を有するものである。

【0019】又、(4) 前記音源位置方向を追尾する手段は、有声音が検出されない音源位置方向の利得を低下させるように前記無指向性マイクロフォンの出力音信号を合成する構成を有するものである。

【0020】又、(5) 前記音源位置方向を追尾する手段は、モニタテレビ装置のスピーカー音信号を検出する手段を有し、前記無指向性マイクロフォンの出力音信号と該モニタテレビ装置のスピーカー音信号との相関を計算して、スピーカー位置を判定し、前記無指向性マイクロフォンから出力されるスピーカー音を抑圧するように、前記無指向性マイクロフォンの出力音信号を合成する構成を有するものである。

【0021】又、(6) 前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とを備えたテレビカメラ装置において、音源位置方向を追尾する手段を優先させて画面を切り出し、話者を強調して表示する映像信号を生成する構成を有するものである。

【0022】又、(7) 前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とを備えたテレビカメラ装置において、移動像を検出して追尾する手段を優先させて画面を切り出し、移動人物を強調して表示する映像信号を生成する構成を有するものである。

【0023】又、(8) 前記音源位置方向を追尾する手段と移動像を検出して追尾する手段とを備え、その両者の追尾手段のいずれかをキー入力により選択する構成を有するものである。

【0024】又、(9) 魚眼又は超広角レンズのカメラと複数の無指向性マイクロフォンとを一体化し、前記魚眼又は超広角レンズを中央部に配置し、前記複数の無指向性マイクロフォンを周辺部に配列して構成したものである。

【0025】

【発明の実施の形態】図1は、本発明の実施の形態のテレビカメラ装置の構成と使用状況を示す図である。図の(A)は、本発明の実施の形態のテレビカメラ装置の平面図、図の(B)は、a1-a2で切断したその断面図である。図の(C)は、本発明の実施の形態のテレビカメラ装置をテレビ会議に使用した場合の使用状況を示す。

【0026】図中、10は全体のテレビカメラ装置、11は魚眼レンズ部、12は無指向性マイクロフォン、13はCCD撮像部、14は会議参加者、15は会議卓、16はモニタテレビ装置である。

【0027】テレビカメラ装置10は、その中央部にカメラの魚眼レンズ部11を配置し、周囲に複数の無指向性マイクロフォン12を、等間隔に配置して構成される。各無指向性マイクロフォン12の出力音声信号は、それぞれ音の特性に応じて遅延量(位相)を変化させて合成することにより、全体として、音の特性に応じて指向性が異なる可変指向性マイクロフォンとして機能する。

【0028】魚眼レンズは約180度の画角を有するため、図1の(C)に示すように、テレビカメラ装置10を、会議参加者が着座する会議卓15の略中央付近に上方を向けて1台設置することにより、会議参加者全員の像を映すことができる。そして、魚眼レンズ部11で撮影した映像の中から、話者又は移動人物等の注目すべき画像を切り出し、切り出した画像の魚眼レンズ歪を補正し、映像信号として出力する。なお、魚眼レンズの代わりに超広角レンズを用いることもできる。

【0029】切り出す画像を、話者又は移動人物の位置識別により自動的に決定して切り替えることにより、旋回台等の可動装置を使用することなく、従って、旋回音の発生がなく、会議参加者は、カメラを意識したりすることなく、自然な雰囲気の中で会議を行うことができる。

【0030】テレビカメラ装置10に配置する無指向性マイクロフォン12は、通常は3乃至4個程であれば充分であるが、より鋭く、精度の高い指向性を得て多人数の会議参加者に対応できるようにするためには、その分より多くの無指向性マイクロフォンを配置すればよい。

【0031】図2は本発明の実施の形態のテレビカメラ装置を用いたテレビ会議システムの構成を示す図である。同図の(A)は画像処理部の構成を示し、図の(B)は音声処理部の構成を示している。2A-1は魚眼レンズカメラ、2A-2は人物位置検出部、2A-3は人物位置決定部、2A-4は画像切り出し・歪み補正部、2A-5は画像通信部(コーデック)、2A-6は遠隔の相手方のテレビ会議システムに接続される回線、2A-7は相手方の会議場の映像を映すモニタテレビ装置である。

【0032】2B-1はマイクロフォンバンク及び音声信号処理部、2B-2は有声音検出部、2B-3は話者位置計算部、2B-4は雑音源位置計算部、2B-5はスピーカー音検出部、2B-6はスピーカー位置計算部、2B-7はマイクロフォン指向性決定部、2B-8は指向性制御部、2B-9は音声通信部(コーデック)、2B-10は相手方のテレビ会議システムに接続される回線、2B-11は音質改善部、2B-12は相手方の会議場の音声を放音するスピーカーである。

【0033】魚眼レンズカメラ2A-1で撮影された画像信号は、人物位置検出部2A-2と画像切り出し・歪み補正部2A-4とに出力され、人物位置検出部2A-2は、移動人物の像及びその位置を検出し、該位置情報を人物位置決定部2A-3に送出する。

【0034】人物位置決定部2A-3は、人物位置検出部2A-2から出力された人物位置情報と、後述する話者位置計算部2B-3から出力される話者位置情報とを基に、魚眼レンズで捉えた画像の中から切り出して映し出す人物位置領域を、設定された優先順序に従って決定する。

【0035】画像切り出し・歪み補正部2A-4は、人物位置決定部2A-3から出力された人物位置領域情報を基に、魚眼レンズの画像の中から当該人物像の画像部分を切り出し、該画像部分の魚眼レンズによる歪みを画像変換操作により補正して、歪みの無い通常の画像の映像信号を、画像通信部2A-5に送出する。

【0036】画像通信部2A-5は、画像切り出し・歪み補正部2A-4から送出された映像信号を伝送方式に適合した符号形式により符号化し、回線2A-6を介して、相手方のテレビ会議システムに送信する。又、画像通信部2A-5は、相手方から回線2A-6を介して受信される相手方の映像符号信号を復号化し、モニタテレビ装置2A-7に相手方の映像信号を送出し、モニタテレビ装置2A-7は、該映像信号により相手方の映像を映し出す。

【0037】図3は前述の人物位置検出部2A-2の構成を示す図である。同図において、3-1はA/D変換器、3-2はフレームメモリ、3-3はフレーム間差分検出部、3-4はフレーム内差分検出部、3-5は移動像輪郭抽出部、3-6は移動人物位置算出部である。

【0038】魚眼レンズカメラ2A-1により撮像した映像は、A/D変換器3-1によりデジタル信号に変換され、フレームメモリ3-2にフレーム単位に記憶され、フレーム間差分検出部3-3により前フレームとの差分が検出され、移動した像の部分が検出される。そして、フレーム内差分検出部3-4により1フレーム内の隣接する画像信号間の差分が検出され、像の輪郭部分が検出される。

【0039】移動像輪郭抽出部3-5は、前記フレーム間差分検出部3-3とフレーム内差分検出部3-4との

出力を結合することにより、移動像の輪郭部分を抽出する。移動人物位置算出部3-6は、移動像輪郭抽出部3-5から出力される移動像の輪郭部分の位置を算出し、その位置情報を図2の人物位置決定部2A-3に出力する。

【0040】図2の人物位置決定部2A-3は、該移動人物の位置情報又は後述する話者位置情報を基に映し出す人物位置領域を決定し、該人物位置領域の情報を、画像切り出し・歪み補正部2A-4に送出する。ここで、画像切り出し・歪み補正部2A-4における魚眼レンズ画像からの表示画像の切り出しと歪み補正の原理について説明する。

【0041】図4は魚眼レンズの画像の切り出しと歪み補正の原理説明図である。同図において、4-1は仮想の画像フレーム、4-2は仮想の半球面、4-3は魚眼レンズの撮像画面である。又、 $x$ 、 $y$ 、 $z$ 軸で示す3次元空間の原点Oに魚眼レンズの中心が置かれ、且つ、魚眼レンズは $z$ 軸の方向に向けて置かれているものとする。

【0042】又、仮想の半球面4-2は、魚眼レンズが正射影方式のレンズでその焦点距離が $f$ であるとする、半径 $f$ の仮想的な半球の球面であり、その平面部は $x-y$ 平面上に置かれ、その中心は原点Oの位置に配置される。

【0043】仮想の画像フレーム4-1は、魚眼レンズの位置(原点)から撮影対象物へ向かう視線方向ベクトルDOV(Direction Of View)と直交する仮想的な平面上のフレームで、所定の大きさの枠を持ち、後に説明するように、この枠内の映像が魚眼レンズ画像から切り出されてモニタテレビ装置の画面に表示されることとなる。即ち、モニタテレビ装置の画像表示フレームの枠と同じ大きさの枠となる。

【0044】又、仮想の画像フレーム4-1は、そのフレーム内に視線方向ベクトルDOVと交差する点を原点とする2次元座標軸( $p$ 、 $q$ )を持ち、仮想の画像フレーム4-1内の点はこの座標の成分( $p$ 、 $q$ )によって表される。

【0045】魚眼レンズによって映し出される像(魚眼レンズ画像)は、レンズの射影方式(projection formula)によっていくつかのタイプに分けられるが、以下、正射影(orthographic projection)方式の魚眼レンズについて説明する。

【0046】魚眼レンズの置かれた位置(原点O)から前記仮想の半球面4-2を通して見える像を、そのまま該仮想の半球面4-2に貼り付けたと仮定する。そして、その半球面4-2上に貼り付けた3次元空間の像を、図の $x$ 軸及び $y$ 軸からなる2次元空間の平面上に、該平面に垂直に( $z$ 方向から原点方向に)押し潰して貼り付けた像が、正射影方式の魚眼レンズの画像である。

【0047】なお、実際には、魚眼レンズの画像は上下

左右逆転した位置に像を結ぶが、像の歪曲の相対的な位置ずれの関係は変わらないので、説明を簡略化するため前述したように結像するものとする。

【0048】前記半球面4-2上の像を垂直に押し潰した2次元空間の平面が魚眼レンズ撮像画面4-3であり、該魚眼レンズ撮像画面4-3の円の内側の画像が魚眼レンズで撮影された画像である。外側の四角の枠は、CCD撮像装置から得られる全体の撮像画面の外枠である。

【0049】魚眼レンズの画像は歪曲しているため、魚眼レンズカメラで撮影した像を表示するには、その一部を切り出して正常な画像に補正する必要がある。そこで、表示するフレームに対応するフレームを仮想の画像フレーム4-1として想定する。

【0050】魚眼レンズが向いているz軸の方向が、鉛直線の上方向の方向であるとし、前記視線方向ベクトルDOVがz軸と成す頂点角(天頂角)を $\phi$ 、水平面の基準軸(x軸又はy軸)と成す方位角を $\theta$ とする。又、焦点距離(魚眼レンズ画像の円の半径)をf、画像フレームの回転角を $\omega$ 、拡大率をmとする。拡大率mは、魚眼レンズの位置(原点O)から画像フレーム4-1内の原点までの距離をDとすると、 $m = D/f$ である。

【0051】魚眼レンズ画像平面4-3内の点(x, y)と画像フレーム4-1内の点(p, q)との対応関係は、下記の関係式によって表わされる。

$$x = R(pA - qB + mR \sin \phi \sin \theta) / (p^2 + q^2 + m^2 R^2)^{1/2}$$

$$y = R(pC - qD - mR \sin \phi \sin \theta) / (p^2 + q^2 + m^2 R^2)^{1/2}$$

$$A = \cos \omega \cos \theta - \sin \omega \sin \theta \cos \phi$$

$$B = \sin \omega \cos \theta + \cos \omega \sin \theta \cos \phi$$

$$C = \cos \omega \sin \theta + \sin \omega \cos \theta \cos \phi$$

$$D = \sin \omega \sin \theta - \cos \omega \cos \theta \cos \phi$$

【0052】上記の関係式を利用して魚眼レンズ画像の歪みを補正し、元の画像を復元してモニタテレビ装置に表示させることができる。即ち、先ず前記画像フレーム4-1を定位させ、モニタ表示する領域(切り出し領域)を定める。この定位操作は、画像フレームの頂点角 $\phi$ 、方位角 $\theta$ 、回転角 $\omega$ 及び拡大率mを、前述の人物位置決定部2A-3から、画像切り出し・歪み補正部2A-4に与えることによって設定される。

【0053】画像切り出し・歪み補正部2A-4は、画像フレーム4-1内の点(p, q)に対応する魚眼レンズ画像平面4-3内の点(x, y)を、前記の関係式により求め、その点の色情報信号を、魚眼画像を記憶した映像フレームメモリ(図示省略)から読み出して、画像フレーム4-1内の点(p, q)に対応するアドレスを有する出力用の映像メモリ(図示省略)に該色情報信号を書き込む。

【0054】魚眼画像を記憶した映像フレームメモリか

ら、画像フレーム4-1の点に対応するアドレスを有する出力用の映像メモリへの色情報転送操作を、画像フレーム4-1内の全ての点(p, q)について行い、出力用の映像メモリから順次、色情報信号を読み出だして出力することにより、画像切り出し・歪み補正部2A-4は、原点Oから仮想の画像フレーム4-1の枠を通して見た実物象の歪みのない映像信号を生成することができる。(米国特許第5,185,667号明細書又は先に出願した特願平10-62531号の明細書等参照)

【0055】次に、図2の(B)を参照して、音声による人物追尾について説明する。複数の無指向性マイクロフォン12を配列したマイクロフォンバンクの出力信号を、音声信号処理部2B-1によりデジタル信号に変換し、有声音検出部2B-2により有声音(母音)が含まれているかどうかを検出し、有声音を含む音声信号に対しては話者位置計算部2B-3により、マイクロフォンバンクの出力信号の位相差から話者位置を計算し、話者位置の情報を人物位置決定部2A-3とマイクロフォン指向性決定部2B-7とに送出する。

【0056】又、有声音を含まない人音信号に対しては雑音源位置計算部2B-4により、同様にその位相差から雑音源の位置を計算し、雑音源位置の情報を、マイクロフォン指向性決定部2B-7に送出する。なお、人音信号に有声音が含まれているかどうかは、線形予測符号化(LPC)のフィルタ係数やケプストラム分析の技術を適用して検出することができる。

【0057】更に、相手方から回線2B-10を介して送信された音声、スピーカー2B-12により放音され、その音声がマイクロフォンバンクに入力されて相手方にエコーとして送出されるのを防止するため、スピーカー音検出部2B-5は、スピーカー2B-12に入力される音声信号とマイクロフォンバンクから入力される音声信号との相関量を検出し、スピーカー位置計算部2B-6は該相関量により、スピーカー2B-12の位置方向を検出し、その位置情報をマイクロフォン指向性決定部2B-7に出力する。

【0058】マイクロフォン指向性決定部2B-7は、話者位置情報、雑音源位置情報及びスピーカー位置情報を基に、各音源に対するマイクロフォンバンクの指向性を決定する。指向性制御部2B-8は、マイクロフォン指向性決定部2B-7で決定された指向性に従って、各音源毎に、各無指向性マイクロフォンの出力信号に異なる位相を与えて合成することにより、音源毎に異なる指向性を与える。

【0059】その結果、有声音を含む話者方向の音源に対しては利得の高い指向性を生じさせ、常時入力されるような雑音源の方向及びモニタテレビ装置のスピーカーの方向に対しては低利得となる指向性を生じさせることにより、音声に対しては高感度で、かつ、雑音やスピーカー音に対しては低感度となり、雑音やスピーカー音を

抑圧し、エコーの発生を低減することができる。

【0060】図5は本発明の実施の形態の雑音を抑圧する指向性を生成する構成の説明図である。同図において、M0乃至M3は無指向性マイクロフォン、5-1はA/D変換器、5-2は相互相関計算部、5-3は音源位置計算部、5-4は話者方向/特性記憶部、5-5はフィルタ及びディレイ部、5-6は音声合成部、5-7はレベル調整/適応フィルタ部、5-8は加算部である。

【0061】マイクロフォンバンクの無指向性マイクロフォンM0乃至M3からの出力信号は、A/D変換器5-1によりA/D変換され、相互相関計算部5-2に入力される。相互相関計算部5-2は、無指向性マイクロフォンM0からの入力信号に対する他の無指向性マイクロフォンM1、M2、M3からの入力信号の相互の相関量と相互位相差とを求め、それらを音源位置計算部5-3に送出する。

【0062】音源位置計算部5-3は、相関量と相互位相差とにより音源位置を計算し、該音源位置の方向を負の指向性（抑圧特性）とするため、各無指向性マイクロフォンM0乃至M3からの入力信号に対して与える相互のディレイ値（位相量）を計算し、その値をフィルタ及びディレイ部5-5に送出する。なお、相互相関計算部5-2と音源位置計算部5-3とは、図2の雑音源位置計算部2B-4とマイクロフォン指向性決定部2B-8に対応する。

【0063】フィルタ及びディレイ部5-5は、音源位置計算部5-3から送出されたディレイ値に応じて、各無指向性マイクロフォンM0乃至M3からの入力信号にディレイを与えて位相を揃え、音声合成部5-6はそれらを合成し、レベル調整/適応フィルタ部5-7は、合成信号のレベルをマイクロフォンM0からの入力信号のレベルに相対的に合わせ、加算部5-8は、レベル調整/適応フィルタ部5-7からの反転出力信号を、マイクロフォンM0からの入力信号に加算して互いに相殺させ、音源に対して負の指向性（抑圧特性）を生成させる。

【0064】従って、加算部5-8から出力される信号は、雑音源等からの不要な音声帯域信号に対して抑圧されたものとなる。なお、図5のフィルタ及びディレイ部5-5と音声合成部5-6とレベル調整/適応フィルタ部5-7と加算部5-8とは、図2の指向性制御部2B-8に対応する。

【0065】又、レベル調整/フィルタ部5-7には、加算部5-8の雑音抑圧音出力のレベルを最小とするような適応フィルタを用いることにより、より精度のよい雑音抑圧効果が得られる。又、雑音源が複数存在する場合には、各雑音源に対してレベルの大きい雑音源から順次同様の処理を行うことにより、複数の雑音源に対して同時に抑圧することができる。

【0066】図6は本発明の実施の形態の抑圧及び強調指向性を生成する構成の説明図である。この構成は、音声検出部を含み、雑音に対して抑圧指向性を生成し、音声に対して強調指向性を生成する構成を示している。同図において、6-1は有声音検出部であり、他の構成は図5に示した構成と同様であるので、同一の構成要素には同一の符号を付し、重複した説明は省略する。

【0067】無指向性マイクロフォンM0からの出力信号を有声音検出部6-1に入力し、有声音検出部6-1は、入力信号に有声音が含まれるかどうかを検出し、その検出結果を音源位置計算部5-3に送出する。なお、有声音検出部6-1は図2に示した有声音検出部2B-2に相当する。

【0068】有声音が含まれるとき、音源位置計算部5-3は、発言者が発声した音声であると判定し、各無指向性マイクロフォンM0乃至M3の相互位相差及び相関量を基に、フィルタ及びディレイ部5-5に対して、その方向の指向性を強調するようにディレイを設定する。

【0069】又、各無指向性マイクロフォンM0乃至M3の相互位相差及び相関量により話者の位置（音源位置）を計算し、該話者位置（音源位置）を話者方向/特性記憶部5-4に送出する。話者方向/特性記憶部5-4は、該話者位置の方向と音源の特性（話者音声）を記憶すると共に、話者位置情報（音源位置情報）を前述の図2の人物位置決定部2A-3に送出する。

【0070】有声音検出部6-1において、有声音が検出されない場合は、雑音と判定し、音源位置計算部5-3は、その音源からの合成信号のエネルギーが最少となるように、フィルタ及びディレイ部5-5のディレイ値（指向性パラメータ）を設定する。なお、この設定の手法は、図5の構成で説明した手法と同様である。

【0071】図7は本発明の実施の形態のスピーカ音抑圧指向性を生成する構成の説明図である。同図において、7-1はスピーカ音検出部である。他の構成は図5に示した構成と同様であるので、同一の構成要素には同一の符号を付し、重複した説明は省略する。

【0072】スピーカ音検出部7-1は、図2に示したスピーカ音検出部2B-5に相当し、モニタテレビ装置のスピーカへの入力信号を参照信号として使用する。スピーカ音検出部7-1において、所定レベル以上のスピーカへの入力信号が検出されると、相互相関計算部5-2において、各無指向性マイクロフォンM0乃至M3とスピーカ音との相互位相差及び相関量を計算し、音源位置計算部5-3は、スピーカ音の抑圧指向性パラメータを算出する。

【0073】そして、雑音抑制の制御の場合と同様に、フィルタ及びディレイ部5-5に対して、スピーカ音が抑圧されるようにディレイを与え、音声合成部5-6はスピーカ音に対しては相互に相殺するように合成する。

【0074】図8は本発明の実施の形態の話者音声と雑音とスピーカー音とを合成する構成の説明図である。同図において、8-1、8-2は合成部、8-3、8-4はフィルタ部である。前述の図5乃至図7の構成で説明したように、話者音声と雑音とスピーカー音とは、異なる特性の音源として区別され、それぞれの音源に対して強調又は抑圧の指向性を与えた後、抽出されるが、該抽出音のうち、話者音声は、合成部8-1、8-2により強調され、雑音はフィルタ部8-3を介した後、合成部8-1により抑圧され、スピーカー音はフィルタ部8-4を介した後、合成部8-2により抑圧され、最終的な合成音が合成部8-2から生成される。

【0075】図9は本発明の実施の形態の話者音声と雑音とスピーカー音とに対する指向性を示す図である。9-1は話者音源、9-2は雑音源、9-3はスピーカー音源、9-4は話者音源に対する指向性、9-5は雑音源に対する指向性、9-6はスピーカー音源に対する指向性、9-7はテレビカメラ装置である。

【0076】前述の図5乃至図7に示す構成により得られる抽出音を、図8に示す構成により合成して出力音を得ることとなる。この出力音は、話者音声に対しては指向性が強く、雑音及びスピーカー音に対しては弱い指向性となる。従って、雑音が低減され、品質の良い話者音声生成されると共に、スピーカー音の回り込みが削減され、エコーキャンセラ又はエコーサプレッサを設けたのと同様に機能し、且つ、本発明のスピーカー音の回り込み削減は、他の音源に対する指向性生成のための構成を兼用して行うことができるため、エコーキャンセラ又はエコーサプレッサを別途設けた場合に比べてはるかに少ない演算処理量で同等の機能を果たすことができる。

【0077】図10は本発明の実施の形態の変向指向性マイクロフォンの指向性制御のフローチャートである。ステップ10-1においてテレビカメラ装置の音声処理部の各部の初期設定を行い、ステップ10-2において監視ループ処理を行い、ステップ10-3においてスピーカー音の検出判定を行い、スピーカー音が検出された場合はステップ10-4において各マイクロフォン入力音とスピーカー音との相互相関を計算し、ステップ10-5において相関パラメータ（位相差、相関量）を計算し、ステップ10-6において、指向性が減少するように、マイクロフォン毎の出力にディレイ及び係数を与えて合成（加減算）し、ステップ10-3に戻る。

【0078】スピーカー音が検出されない場合は、ステップ10-7において有声音の音量が大であるか判定し、大であるときは話者方向を追尾するよう、ステップ10-8において各マイクロフォン入力音間の相互相関を計算し、ステップ10-9において相関パラメータ（位相差、相関量）を計算し、ステップ10-6において指向性が増大するように、マイクロフォン毎の出力にディレイ及び係数を与えて合成（加減算）し、ステップ

10-3に戻る。

【0079】有声音の音量が小であるときは、雑音を打ち消すよう、ステップ10-10において各マイクロフォン入力音間の相互相関を計算し、ステップ10-11において相関パラメータ（位相差、相関量）を計算し、ステップ10-6において指向性が減少するように、マイクロフォン毎の出力にディレイ及び係数を与えて合成（加減算）し、ステップ10-3に戻る。

【0080】次に、本発明の画像及び音声による統合追尾について説明する。前述したように魚眼画像による移動像の追尾、及び複数の無指向性マイクロフォンの入力音の位相差による話者の追尾の組み合わせにより、移動人物の存在方向の検出とその追尾、及び話者の方向の検出とその追尾が可能になる。

【0081】魚眼レンズカメラによる全天周映像の中から、一部の人物を切り出す操作は、複数の人物に対して同時に並行して実行することができ、その切り出しにおいて、移動人物又は話者の検出結果は、互いに独立に採用することが可能であり、外部からの設定によって動人物追尾優先、話者追尾優先、等の選択を行うことが可能となる。

【0082】前述の図2に示した人物位置決定部2A-3は、人物位置検出部2A-2又は話者位置計算部2B-3から得られる移動人物位置情報又は話者位置情報に基づいて、設定された優先順序に従って、移動像の追尾又は音源位置方向の追尾を行う。又、移動像の追尾或いは音源位置方向の追尾を、キー入力により選択するように構成することもできる。

【0083】図11は、本発明の実施の形態の送出画像の例を示す図である。図の(A)は4人の人物に対して同時に並行して切り出し操作を実行し、4画面の合成映像としたものである。それぞれの人物が会話をせずに移動するような場合には、それぞれの個々の人物追尾機能が働き、常時分割された映像の中に、人物が表示される。

【0084】又、人物が静止した状態で会話を続けるような場合には、音声追尾による位置決め精度の向上により、人物の顔部分の特定が可能になり、図(B)の右上画面又は図の(C)の左下画面のように、ズームアップを行うことができる。当然、会話を行う人物が変わったり、複数存在する場合には、それに合わせて人物をズームアップすることが可能になる。このように、話者に対してズームングを行うこと等で、従来に比べ、よりアクティビティのある映像を、人が操作を行うことなく確実にかつ迅速に行うことができる。

【0085】

【発明の効果】以上説明したように、本発明によれば、魚眼又は超広角レンズのカメラと、複数の無指向性マイクロフォンを配列した変向指向性マイクロフォンとにより、画像及び音声を全て収集し、その中から話者又は移



動人物等の画像を切り出して表示する構成としたことにより、可動部分が全くなく、会議参加者に違和感を与えないコンパクトでシンプルなテレビカメラ装置を構成することができる。

【0086】又、複数の無指向性マイクロフォンによる音源位置方向の特定において、有声音の検出及びスピーカ音との相関の検出を行い、音源の音質に応じて異なる指向性を生成することにより、雑音やエコーの少ない高品質の臨場感のあるテレビ会議を実現することができる。

【0087】又、魚眼又は超広角レンズのカメラによる移動人物追尾と、可変指向性マイクロフォンによる話者追尾の機能を備えたことにより、移動せずに発言する人物の画像切り出し、及び音声を発生せずに移動する人物の画像切り出しを、安定的に高速且つ高精度で行うことができ、アクティビティのある画像を生成することができる。

【0088】又、監視用のテレビカメラ装置等として使用した場合に、逆光や照明不足などのため移動像の切り出しが難しい場合でも、音声の発生方向を検知することにより、安定した人物の切り出しと追尾が可能となる等、使用状況に適応した切り出し処理を設定することができる。

【図面の簡単な説明】

【図1】本発明の実施の形態のテレビカメラ装置の構成と使用状況を示す図である。

【図2】本発明の実施の形態のテレビカメラ装置を用いたテレビ会議システムの構成を示す図である。

【図3】本発明の実施の形態のテレビカメラ装置の人物

位置検出部の構成を示す図である。

【図4】魚眼レンズの画像の切り出しと歪み補正の原理説明図である。

【図5】本発明の実施の形態の雑音を抑圧する指向性を生成する構成の説明図である。

【図6】本発明の実施の形態の抑圧及び強調指向性を生成する構成の説明図である。

【図7】本発明の実施の形態のスピーカ音抑圧指向性を生成する構成の説明図である。

【図8】本発明の実施の形態の話者音声と雑音とスピーカ音とを合成する構成の説明図である。

【図9】本発明の実施の形態の話者音声と雑音とスピーカ音とに対する指向性を示す図である。

【図10】本発明の実施の形態の可変指向性マイクロフォンの指向性制御のフローチャートである。

【図11】本発明の実施の形態の送出画像の例を示す図である。

【図12】従来のテレビ会議用のカメラ旋回システムを示す図である。

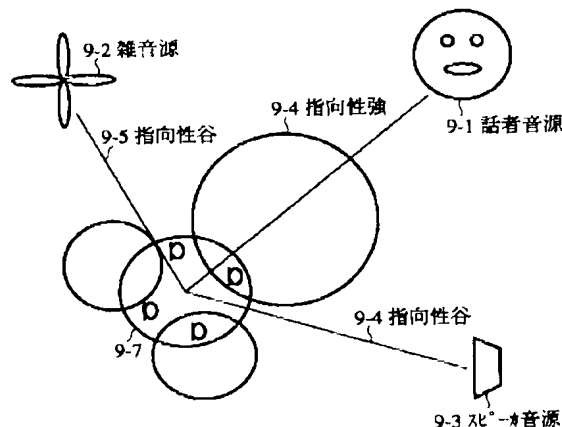
【図13】従来の話者位置を表示するテレビ会議システムを示す図である。

【符号の説明】

- 10 テレビカメラ装置
- 11 魚眼レンズ部
- 12 無指向性マイクロフォン
- 13 CCD撮像部
- 14 会議参加者
- 15 会議卓
- 16 モニタテレビ装置

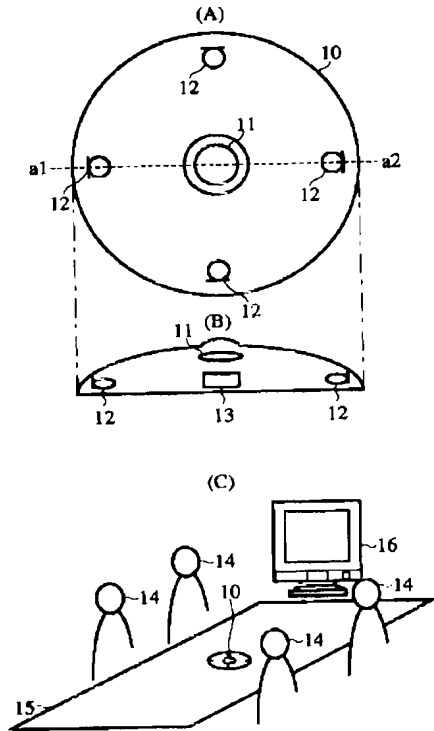
【図9】

本発明の実施の形態の話者音声と  
雑音とスピーカ音とに対する指向性を示す図



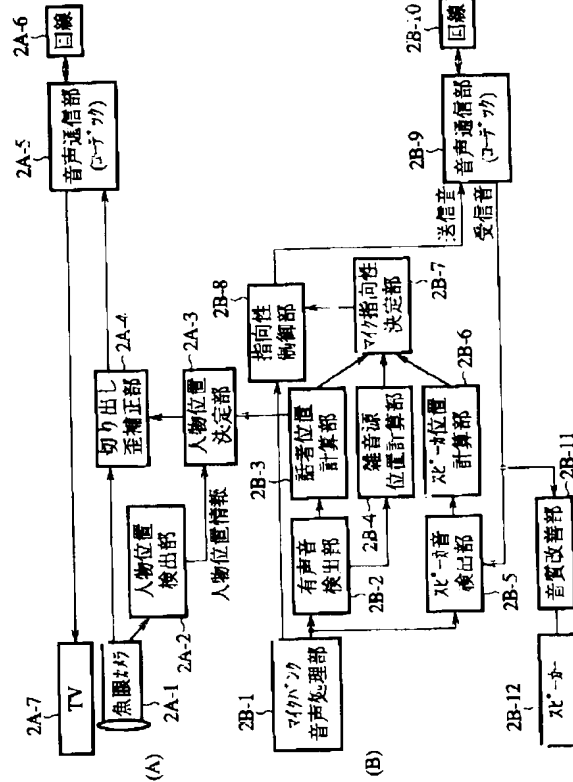
【図1】

本発明の実施の形態の  
テレビカメラ装置の構成と使用状況を示す図



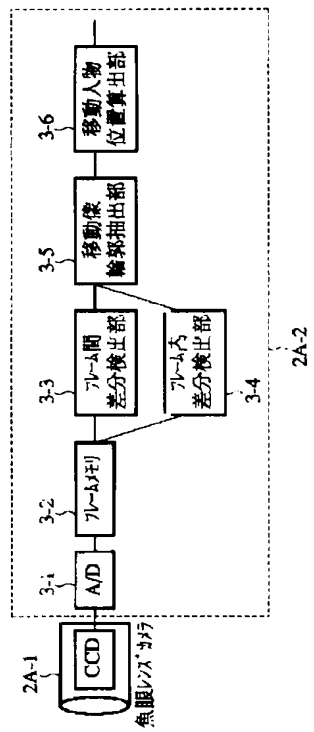
【図2】

本発明の実施の形態のテレビカメラ装置を  
用いたテレビ会議システムの構成を示す図



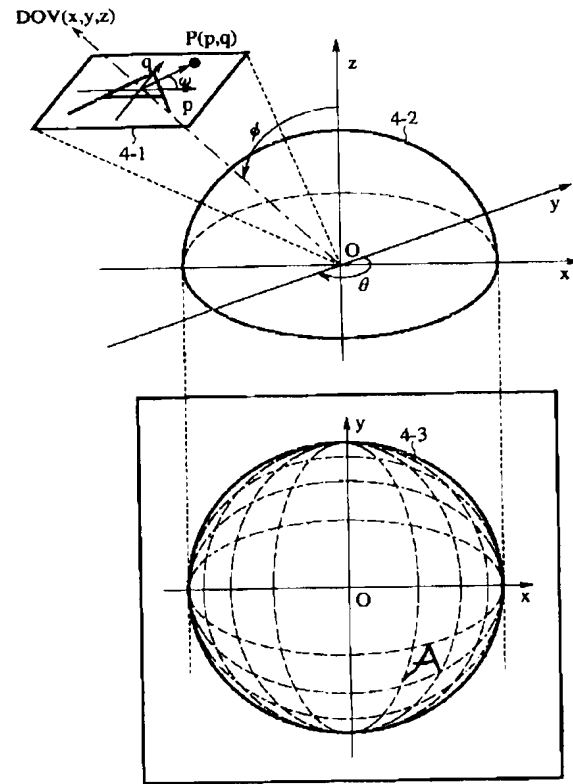
【図3】

本発明の実施の形態のテレビカメラ装置の  
人物位置検出部の構成を示す図



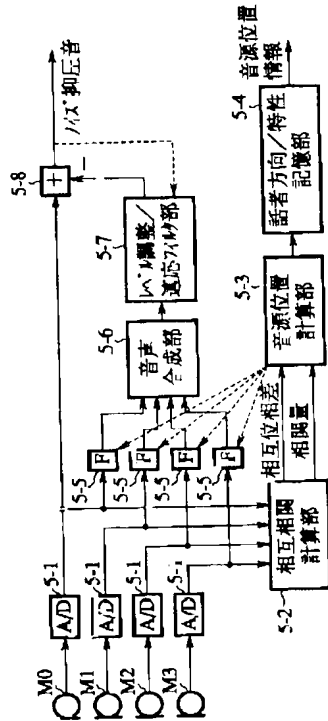
【図4】

魚眼レンズの画像の切り出しと歪み補正の原理説明図



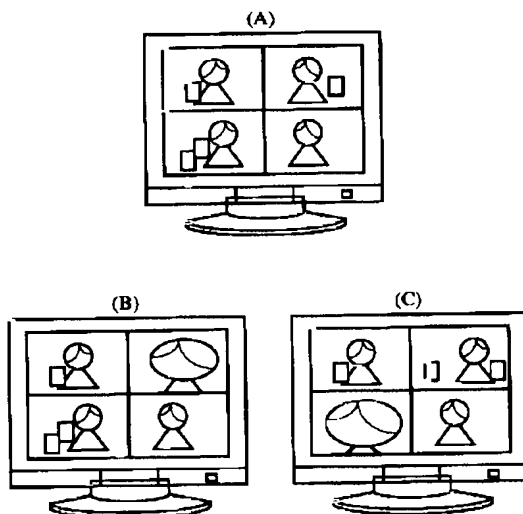
【図5】

本発明の実施の形態の雑音を抑制する指向性を生成する構成の説明図



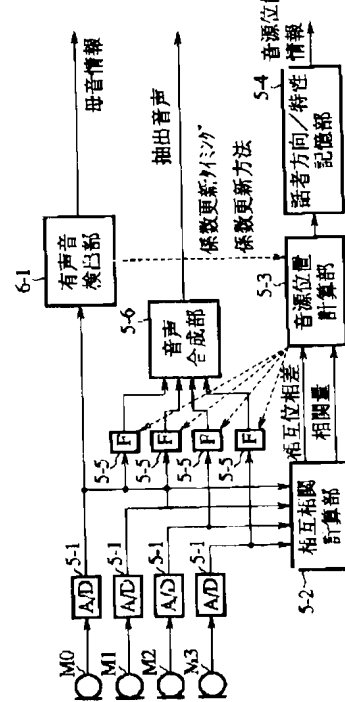
【図11】

本発明の実施の形態の送出画像の例を示す図



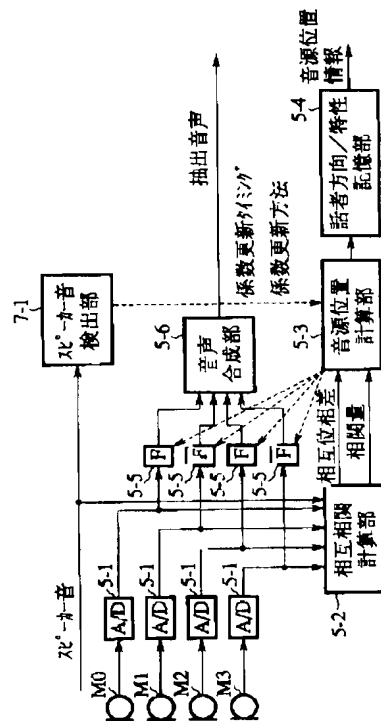
【図6】

本発明の実施の形態の抑圧及び強調指向性を生成する構成の説明図



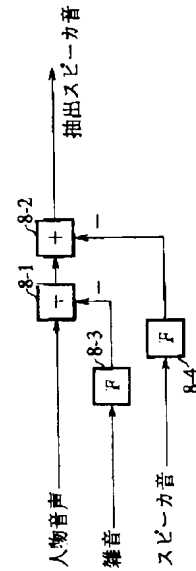
【図 7】

本発明の実施の形態の  
スピーカ音抑圧指向性を生成する構成の説明図



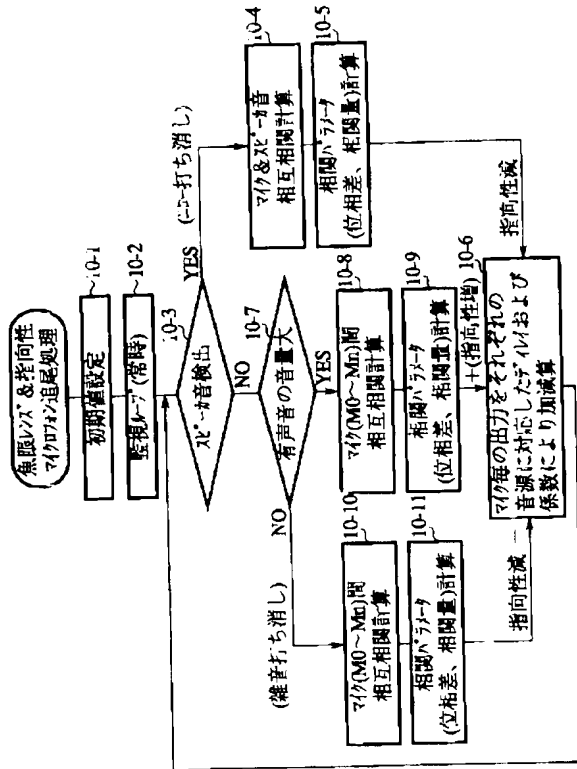
【図 8】

本発明の実施の形態の話者音声と雑音と  
スピーカ音を合成する構成の説明図



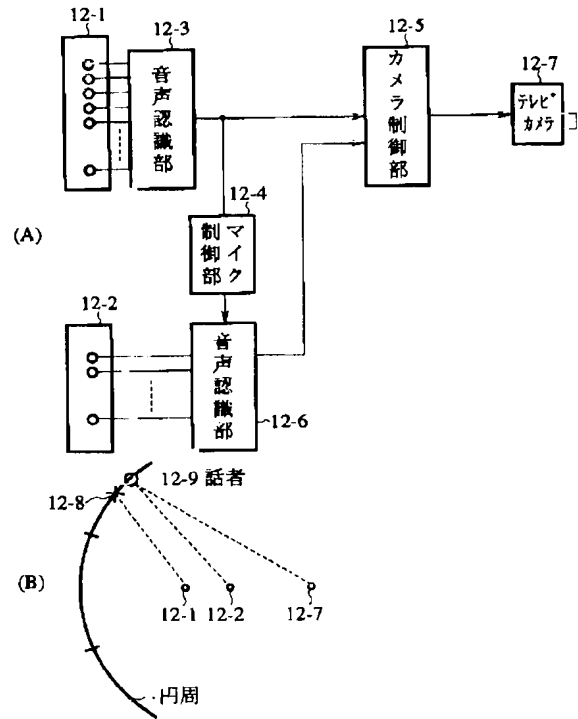
【図10】

本発明の実施の形態の変指向性マイクロフォンの  
指向性制御のフローチャート



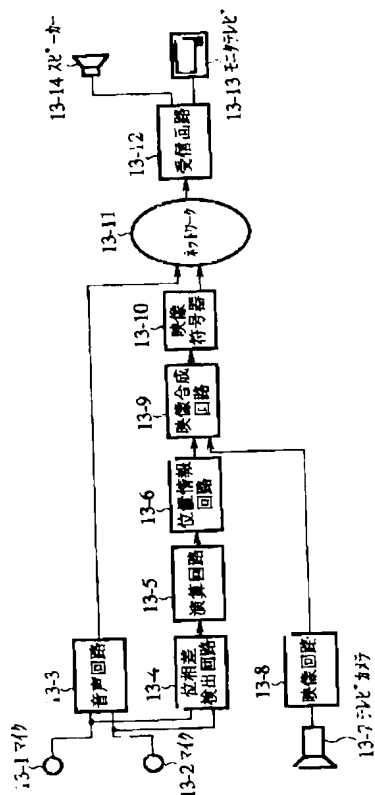
【図12】

従来のテレビ会議用のカメラ旋回システムを示す図



【図 1 3】

従来の話者位置を表示するテレビ会議システムを示す図



フロントページの続き

(51)Int. Cl.<sup>6</sup>H 0 4 N 5/268  
7/15

識別記号

F I

H 0 4 N 7/15

G 0 6 F 15/62

15/64

15/66

3 8 0

3 4 0 B

4 7 0 A